Data Resource Profile

Data Resource Profile: The Russia Longitudinal Monitoring Survey—Higher School of Economics (RLMS-HSE) Phase II: Monitoring the Economic and Health Situation in Russia, 1994–2013

Polina Kozyreva,[1] Mikhail Kosolapov[2] and Barry M Popkin[3]*

[1]Center for Longitudinal Studies, National Research University Higher School of Economics and Institute of Sociology, Russian Academy of Sciences, Moscow, Russia, [2]Institute of Sociology, Russian Academy of Sciences and Independent Research Center "Demoscope", and National Research University Higher School of Economics, Moscow, Russia and [3]Carolina Population Center and Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

*Corresponding author. Carolina Population Center, University of North Carolina, 137 East Franklin St, Chapel Hill, NC 27516, USA. E-mail: popkin@unc.edu

Why was the data resource set up?

The Russian Longitudinal Monitoring Survey (RLMS) was initially created by the G-7 countries in 1992 as a way to obtain objective nationally representative data on the social, health and economic situation in Russia. It was established to mirror a multipurpose survey—the China Health and Nutrition Survey[1]—and provide in-depth reliable raw data on Russia, accessible for the first time to both Russian and global scholars and institutions. This was instituted in the period following January 1992, when the Russian Federation introduced a series of sweeping economic reforms, including eliminating most food and reducing fuel and other subsidies, using freely fluctuating market prices, privatizing many state enterprises and working to create a growing private sector with private land ownership.

The RLMS was created because the existing data, including a Family Budget Survey, were deemed unreliable, and adequate dietary, anthropometric and various other health-related behaviours were not measured in a nationally representative manner. These problems led to the initial Phase I survey of four rounds (I–IV) which was discontinued and is described in Supplement 1 (available as Supplementary data at *IJE* online). This was the first nationally representative random sample of economic and health data ever collected in Russia, with all earlier sampling based on quotas from enterprises and other organizations.

The ongoing longitudinal survey began in 1994 with the Phase II survey. In 2010, the Higher School of Economics (HSE) brought a number of the senior RLMS scholars onto its faculty and began to provide funding for the RLMS. Supplementary funding for subsequent nutrition and health-related data came from the University of North Carolina. At this time a decision was made to change the name to the RLMS-HSE.

Data resource basics for the phase II survey

Sample design Phase II

The target sample size was set at 4 000 households. A multistage probability sample of households was

employed to get a nationally representative sample for the Russian Federation. First, a list of 1850 consolidated raions (administrative-territorial districts), containing 95.6% of the population, was created to serve as primary sampling units (PSUs). These were allocated into 38 strata based largely on geographical factors and level of urbanization, but also based on ethnicity where there was salient variability. Three very large population units were selected with certainty: Moscow city, Moscow Oblast and St Petersburg city constituted self-representing (SR) strata. The remaining non self-representing raions (NSR) were allocated to 35 equal-sized strata. The total of 98 PSUs were selected: 63 PSUs in three self-representing strata and 35 PSUs in the rest non-representative strata. In urban areas of the selected PSUs, secondary sampling units (SSUs) were defined by the boundaries of census enumeration districts. In rural areas, villages were compiled to serve as SSUs.

This was designed as an annual survey. Two years were missed, 1997 and 1999, due to funding lapses between 1994 and 2014. The sample is described in more detail in Supplement 2, Phase II (available as Supplementary data at *IJE* online) and on the RLMS-HSE websites [http://www.cpc.unc.edu/projects/rlms-hse/project/sampling].

In both urban and rural substrata, interviewers were required to visit each selected dwelling up to three times to secure the interviews. They were not allowed to make substitutions of any sort. 'Household' was defined as a group of people who live together in a given domicile and share common income and expenditures. Households were also defined to include unmarried children, 18 years of age or younger, who were temporarily residing outside the domicile at the time of the survey.

The interviewer then conducted individual interviews with as many household members aged 14 and older as possible, acquiring data about their individual activities and health. Data for children aged 13 and younger were obtained from adults in the household. This provided a probability sample of Russian individuals without special weighting at baseline.

Nationally representative sample

The sample frame was essentially based on dwellings. In conducting rounds VI–XXII, interviewers in both urban and rural areas attempted to conduct interviews in the same dwellings that fell into the first round of Phase II, round V sample. They returned to each round V dwelling even if the household had refused to participate during previous rounds, and even if they found out that the household whom they interviewed in previous rounds had moved to a new dwelling before the interview. In Moscow and St Petersburg, where the greatest non-response and accordingly the greatest attrition rates of the sample were observed, the sample was replenished several times and this was undertaken once in a few other cities. Figure 1 provides the dynamics of sample sizes of Phase II and describes the series of replenishments that occurred over time to get to the final RLMS-HSE sample size from the round XXII in 2013.
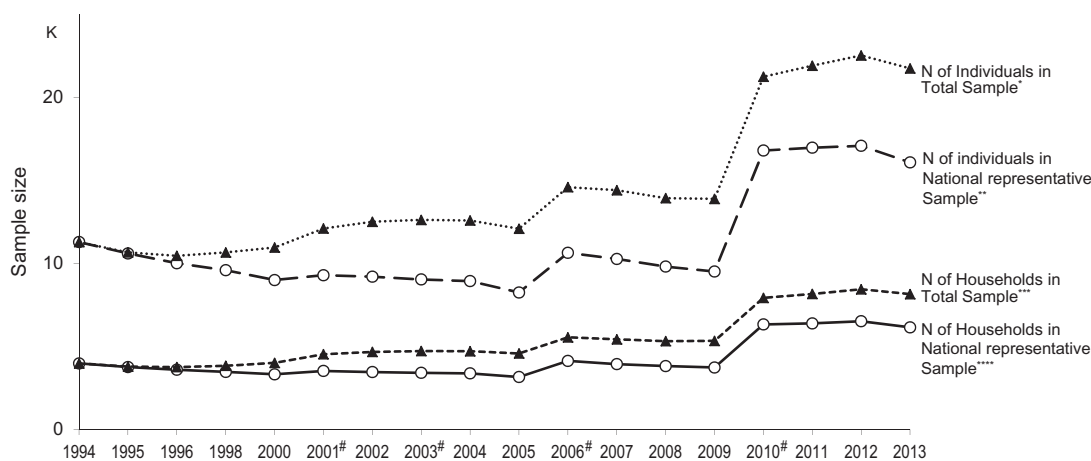

Figure 1. The dynamics of sample sizes Phase II RLMS-HSE2001. The nationally representative sample is followed by interviewing households and individuals residing at the addresses of 1994 sample and addresses of replenishments. The total sample includes in addition the movers (households or individuals who moved to new units for any reason, and were followed). #Replenishments: 2000, replenishment samples in Moscow and St Petersburg, 2003: replenishment of the region within a stratum in 2003 (Novosibirsk region instead of Khanty-Mansiisk region); 2006, replenishment to 1994 sample in most regions; 2010, a 50% increase in sample size following an identical sample selection approach.

*All individuals, participating in a given round, including movers who were followed. **Only individuals residing at the addresses of 1994 sample and addresses of replenishment. ***All households participating in a given round, including movers who were followed. ****Only households residing at the addresses of 1994 sample and addresses of replenishment.

Longitudinal cohort

The original sampling plan did not call for households to be followed if they moved from the round V (1994) sample dwelling unit. Likewise, individual household members who moved away were not to be followed. After round VII (1996), all individuals and households were followed when they moved out of the household units (families, separated, children got married, and so on) to live in the same secondary sampling unit (SSU) or move into one of the PSUs in the sample. This created the current longitudinal cohort. We attempted to find households who moved in the 1994–96 period also.

Multilevel design

An array of contextual economic, demographic, social and built environment infrastructure and related data are collected for each of the smallest sampling units or local communities (essentially SSUs or villages).

In all rounds of Phase II, questionnaires were obtained from over 97% of the individuals listed on the household rosters. The distribution of household size in the sample, within both rural and urban localities, corresponds well to the figures from the Russian census during all rounds of the survey (Supplement 2, Table 3, available as Supplementary data at *IJE* online). Bear in mind that single-member households are excluded from the comparison because the census includes many institutionalized people,

whereas our sample explicitly excludes them. Thus, there is no valid basis for comparison.

The multivariate distribution of the sample by sex, age, education and urban-rural location compares quite well with the corresponding multivariate distributions of the nearest census data (Supplement 2, Tables 4 and 5, available as Supplementary data at *IJE* online). There are usually the differences of only 1–2 percentage points between these distributions. The ethnic composition of the sample throughout all rounds also corresponds to the census figures, having about 86% of Russians, 2.4% of Tatars and 10% of other nationalities.

Response rates

The household response rate in round V (which was the first round of Phase II) exceeded 87.6% (for more detail see Supplement 2, Tables 6 and 7, available as Supplementary data at *IJE* online). Table 1 shows that over half of the households participated in 10 rounds of RLMS-HSE, and for individuals about half participated in eight rounds. This creates a good basis for longitudinal analysis.

The response rates varied across PSUs, depending on the proportion of households in rural areas. Obviously, in Moscow and St Petersburg, respondents and household response rates are substantially lower than in the Russian Federation as a whole and, of course, the whole of Russia without these two cities (Supplement 2, Tables 6 and 7). However, since this situation was expected and has been

Table 1. The duration of participation in the survey (participation rate) for 1994 households and individuals (including separated or moved out) 1994–2013

Rounds participated	Household		Individual	
	Percentage	Cumulative percentage	Percentage	Cumulative percentage
All 18 rounds	26.14	26.14	16.50	16.50
Seventeen rounds	6.59	32.73	6.18	22.68
Sixteen rounds	3.55	36.28	3.80	26.48
Fifteen rounds	2.74	39.02	3.46	29.95
Fourteen rounds	2.74	41.76	2.92	32.87
Thirteen rounds	2.77	44.53	2.69	35.56
Twelve rounds	2.67	47.19	2.95	38.51
Eleven rounds	3.22	50.42	2.96	41.47
Ten rounds	3.12	53.53	2.93	44.40
Nine rounds	2.74	56.28	3.08	47.48
Eight rounds	2.99	59.27	3.21	50.69
Seven rounds	2.92	62.19	3.22	53.91
Six rounds	3.19	65.38	3.90	57.80
Five rounds	3.47	68.86	4.09	61.90
Four rounds	5.69	74.54	6.32	68.22
Three rounds	6.99	81.53	8.97	77.19
Two rounds	6.67	88.20	8.32	85.51
One round	11.80	100.00	14.49	100.00

adjusted in oversampling procedures, the actual proportion of completed household interviews compares well to the proportion of the population in each stratum.

Since the highest non-response rate occurred in Moscow and St Petersburg, the duration of participation in the survey in these two cities was the lowest (Supplement 2, Tables 8–11, available as Supplementary data at *IJE* online).

Attrition rates

One of the most important questions is: 'How misleading would it be to conduct pure panel analysis of households and individuals observed in any set of consecutive rounds?' The obvious problem is that, by definition, pure panel analysis can include only those who continue to reside in the original sample dwelling units and participate in this set of consecutive rounds. To evaluate the possibility of such analysis, it is necessary to calculate attrition rates for any such sequence of rounds. As an example, we present calculations for two most popular types of attrition rates (Supplement 2, Tables 12–14, available as Supplementary data at *IJE* online), namely wave-to-wave and baseline-on-wave attrition rates for individuals and households. For all 18 rounds, only about 29% of households and 19% of individuals continued to participate (1994–2014) but, if we look at the first 10 years, the results were about 60% and 51%, respectively (rounds 1–9) (Supplementary Table 12). Table 2 presents death rates for the initial 1994 participants. Overall, 12.8% have passed away.

Data collected

Throughout the entire set of surveys, very detailed basic household and individual data have been collected. Table 3 details this set of economic, labour force, demographic, education, and related socioeconomic data. The full set of English and Russian survey instruments are available on the two RLMS-HSE websites. The household and individual core socioeconomic data are extremely detailed. They contain classic income and expenditures data on all categories, from weekly food purchases to consumer durables. The demographic data provide a classic triangle of the relationships of each person with each other within the household. The asset data include all sorts of details on household and other assets. The employment information

is in-depth for multiple jobs with detail on type of employment, earnings, hours and ownership status (public, private, joint) and provides the four-digit International Labour Organization occupation code. Both actual and perceived quality of life questions are interspersed.

Health data: for each wave, detailed data on alcohol and smoking were obtained. Health service use data are also collected but not in great detail. For selected rounds, direct measurement of weight, height and waist circumference were obtained (rounds V–XIV and XX). Also one-day 24-h recall dietary data were obtained in these rounds. In only one round were replicates of a second day collected for the sample.[2,3] Nutrient intake levels are reported; however, actual detailed dietary data are not available as the food composition table and data are controlled by a collaborator and were not made available.

There have been attempts to obtain biomarkers; unfortunately, fasting blood or blood spot collection has been impossible as blood samples in any form cannot be taken out of Russia, and it has not been possible to find a laboratory equipped to handle full blood spot assays at reasonable cost and reliability. These data have yet to be collected.

Spatial coordinates

For some time we attempted to use global positioning technology and collect coordinates for all major social and economic and transport and health-related infrastructure as well as household coordinates. Politically this was not feasible until recently, and funding has not been obtained to undertake this collection. However, the survey team is able to provide (at cost) linkages of external data sets to the RLMS-HSE contextual data by using deductive disclosure controls to ensure anonymity of the identification of communities.

Data resource use

Hundreds of English-language publications have arisen from the RLMS-HSE data, authored by scholars globally. In addition, there are thousands of Russian-language publications which are not accessible to most scholars globally. Most of the focus has been on the poverty, economic, social and demographic data. These dietary and socioeconomic data were used to create the Russian poverty line, which established the pension level such that few

Table 2. Percentages of 1994 participants who died between 1994 and 2013

1995	1996	1998	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
0.69	1.05	1.70	1.49	0.88	0.98	0.88	0.74	0.70	0.60	0.49	0.54	0.47	0.43	0.44	0.39	0.34	12.80

Table 3. RLMS-HSE survey components

Round	Year of collection	Core household SES data[a]	Core individual SES data[b]	Time budget	24-h diet/ weight- height-WC	Child care	Abortion/ family planning	Sexual behaviour, confidential
V	1994	X	X	X	X	X	X	
VI	1995	X	X	X	X	X	X	
VII	1996	X	X	X	X	X	X	
VIII	1998	X	X	X	X	X	X	
IX	2000	X	X		X	X	X	
X	2001	X	X		X	X	X	X
XI	2002	X	X		X	X	X	
XII	2003	X	X		X	X	X	X
XIII	2004	X	X		X	X	X	
XIV	2005	X	X		X	X	X	
XV	2006	X	X			X	X	
XVI	2007	X	X			X	X	
XVII	2008	X	X			X	X	
XVIII	2009	X	X			X	X	
XIX	2010–11	X	X			X	X	
XX	2011–12	X	X		X	X	X	
XXI	2012–13	X	X			X	X	
XXII	2013–14	X	X			X	X	

WC, waist circumference.

[a]The core household data collected each year include: household composition/relationships; housing (structure, amenities, privatization, ownership); possession of consumer durables; raising food on private plots; in-depth food, clothes and consumer durables during 3 months, savings, transfer payments, gifts to others, utilities and many other expenditures; income from all wage and non-wage sources by public and private sector status, including transfer payments, gifts, stock market, and drawing down savings; and details on non-payment of wages and losses due to bank closures.

[b]The core individual data (questions on children age < 14, answered by parents): these include place of birth, some migration, language, marital status; work (primary, secondary, entrepreneur, independent, unofficial, unemployment, employment-seeking); years of work experience; willingness to be retrained; four-digit occupational coding according to the International Labour Organization protocol; education (current and past); self-ratings of satisfaction, well-being, poverty, relationship with others; use of medical services and medicines and insurance; childbearing and birth control (including child-bearing and abortion history); plans included are smoking and alcohol in-depth blocks of questions.

pensioners in Russia are in poverty[4,5] and almost none suffer weight loss due to a lack of income.[6] Related to the poverty line has been extensive research on poverty by the World Bank and many scholars globally.[5,7,8]

Alcohol intake has been subject to serious examination by a vast number of scholars.[9] One of the more interesting issues is the skewed distribution with a small proportion of men of all ages consuming about a half-litre of actual alcohol per day.[10–12] The data showed a decreasing prevalence of drinking during this period but an increase in the amount of alcohol consumed by some members of this population, and important cohort effects with older Russians more likely to be drinking excessively.[13] Partly because of the high alcohol intake levels and the stresses of the economic transformation, overall health, life expectancy and mortality have been studied extensively.[14–17] A third topic is abortion, for which the RLMS-HSE results produced much lower estimates than previous research.[18] According to RLMS-HSE data, the abortion rate in 1994 was 56 per 1000 women aged 15–44, with a 95% confidence interval of ± 12 per 1000, an estimate that varies from that advanced by official sources and other studies.

Part of the reason for this difference is that the government listed all miscarriages as induced abortions. In addition, we used the advice of demographers who had studied this issue for years (Professor Barbara Anderson, University of Michigan, and others) to create confidential interviews on this component.

Strengths and weaknesses

The major strengths of the RLMS-HSE are the national representativeness, collection of very high quality sociodemographic and economic data, and the long follow-up. The biggest weaknesses from the health side are the lack of biomarkers and erratic collection of dietary and body composition data based on outside funding availability. And as in all longitudinal surveys the attrition over time should be considered while analyzing the data.

Data resource access

The bulk of the RLMS-HSE data are completely free and available on the RLMS websites in English [http://www.

cpc.unc.edu/projects/rlms-hse/] and Russian and English [www.hse.ru/rlms]. The sexual behaviour data are highly confidential, as are spatial locations of sample recipients. Institutional Review Board approval for each survey has been provided by both the institutional review boards of the University of North Carolina and the Higher School of Economics. Contextual data require also special applications. To link other contextual measures to the RLMS-HSE data, this must be done at cost by contacting the Carolina Population Center.

Supplementary data

Supplementary data are available at *IJE* online.

Acknowledgments

Profile in a nutshell

- The RLMS-HSE was established to create a nationally representative survey to monitor the economic and health impact of the massive set of reforms in the Russian Federation.
- Established in 1992 (Phase I) and, for Phase II (discussed in depth here) in 1994, this annual survey is both nationally representative plus has a longitudinal component. Both collect multipurpose health and economic studies with in-depth individual, household and community contextual data collected

---

in all rounds.
- The 1994 and 2013 samples collected the data from 11290 and 21753 individuals and 3975 and 8149 households, respectively. A multistage sample with 98 primary sampling units (Moscow city, Moscow Oblast and St Petersburg are self-representing) was designed to represent the Russian Federation.
- The major data components are: economic (detailed income, labour force behaviour and expenditures data); demographic/sociological (household structure and age-gender composition, background, education and school behaviour); and health (24-h dietary recall, smoking, drinking activity, body mass index direct measurement).
- Data can be linked to other contextual datasets.
- The bulk of the RLMS-HSE data are completely free and available on the UNC-CPC websites in English [http://www.cpc.unc.edu/projects/rlms-hse/] and the RLMS-HSE website in Russian and in English [http://www.hse.ru/rlms/]. Selected confidentiality forms are required for selected data such as sexual behaviour of adolescents.

Funding

Conflict of interest: None declared.

References

1. Popkin BM, Du S, Zhai F, Zhang B. Cohort Profile: The China Health and Nutrition Survey—monitoring and understanding

socio-economic and health change in China, 1989-2011. *Int J Epidemiol* 2010;**39**:1435–40.

2. Jahns L, Arab L, Carriquiry A, Popkin BM. The use of external within-person variance estimates to adjust nutrient intake distributions over time and across populations. *Public Health Nutr* 2005;**8**:69–76.

3. Jahns L, Carriquiry A, Arab L, Mroz TA, Popkin BM. Within- and between-person variation in nutrient intakes of Russian and US children differs by sex and age. *J Nutr* 2004;**134**:3114–20.

4. Lokshin M, Harris KM, Popkin BM. Single mothers in Russia: household strategies for coping with poverty. *World Dev* 2000;**28**:2183–98.

5. Lokshin M, Popkin BM. The emerging underclass in the Russian Federation: income dynamics. *Econ Dev Cult Change* 1999;**47**:803–29.

6. Stookey JD, Zohoori N, Popkin BM. Nutrition of elderly people in China. *Asia Pac J Clin Nutr* 2000;**9**:243–51.

7. Lokshin M, Ravallion M. Welfare impacts of the 1998 financial crisis in Russia and the response to the public safety net. *Econ Transit* 2000;**8**:269–95.

8. Ravallion M, Lokshin M. Who wants to redistribute Russia's tunnel effect in the 1990s? *J Public Econ* 2000;**76**:87–104.

9. Zohoori N, Mroz TA, Popkin B *et al*. Monitoring the economic transition in the Russian Federation and its implications for the demographic crisis—The Russian Longitudinal Monitoring Survey. *World Dev* 1998;**26**:1977–93.

10. Baltagi BH, Geishecker I. Rational alcohol addiction: evidence from the Russian Longitudinal Monitoring Survey. *Health Econ* 2006;**15**:893–914.

11. Tapilina VS. How much does Russia drink: volume, dynamics and differentiation of alcohol consumption. *Russian Soc Sci Rev* 2007;**48**:79–94.

12. Tekin E. Employment, wages, and alcohol consumption in Russia. *South Econ J* 2004;**71**:397–417.

13. Zohoori N. Recent patterns of alcohol consumption in the Russian elderly, 1992-1996. *Am J Clin Nutr* 1997;**66**:810–14.

14. Andreev EM, McKee M, Shkolnikov VM. Health expectancy in the Russian Federation: a new perspective on the health divide in Europe. *Bull World Health Organ* 2003;**81**:778–87.

15. Cockerham WC. Health lifestyles in Russia. *Soc Sci Med* 2000;**51**:1313–24.

16. Perlman F, Bobak M. Socioeconomic and behavioral determinants of mortality in posttransition Russia: a prospective population study. *Ann Epidemiol* 2008;**18**:92–100.

17. Shkolnikov VM, Andreev EM, Anson J, Mesle F. The peculiar pattern of mortality of Jews in Moscow, 1993–95. *Popul Stud* 2004;**58**:311–29.

18. Entwisle B, Kozyreva P. New estimates of induced abortion in Russia. *Stud Fam Plann* 1997;**28**:14–23.